BARD Final Report IS-4744-14CR

**Population genomics, linkage disequilibrium and association mapping of stripe rust resistance genes in wild emmer wheat, *Triticum turgidum ssp. dicoccoides***

Sela, H. Tel Aviv University
Akhunov, E. Kansas State University KS
Steffenson, B.J. University of Minnesota MN

Project award year: 2014
Three year research project

**Abstract**

The primary goals of this project were: (1) development of a genetically characterized association panel of wild emmer for high resolution analysis of the genetic basis of complex traits; (2) characterization and mapping of genes and QTL for seedling and adult plant resistance to stripe rust in wild emmer populations; (3) characterization of LD patterns along wild emmer chromosomes; (4) elucidation of the multi-locus genetic structure of wild emmer populations and its correlation with geo-climatic variables at the collection sites.

Introduction

In recent years, Stripe (yellow) rust (Yr) caused by Puccinia striiformis f. sp. *tritici* (PST) has become a major threat to wheat crops in many parts of the world. New races have overcome most of the known resistances. It is essential, therefore, that the search for new genes will continue, followed by their mapping by molecular markers and introgression into the elite varieties by marker-assisted selection (MAS). The reservoir of genes for disease and pest resistance in wild emmer wheat (*Triticum dicoccoides*) is an important resource that must be made available to wheat breeders. The majority of resistance genes that were introgressed so far in cultivated wheat are resistance (R) genes. These genes, though confering near-immunity from the seedling stage, are often overcome by the pathogen in a short period after being deployed over vast production areas. On the other hand, adult-plant resistance (APR) is usually more durable since it is, in many cases, polygenic and confers partial resistance that may put less selective pressure on the pathogen.

In this project, we have screened a collection of 480 wild emmer accessions originating from Israel for APR and seedling resistance to PST. Seedling resistance was tested against one Israeli and 3 North American PST isolates. APR was tested on accessions that did not have seedling resistance. The APR screen was conducted in two fields in Israel and in one field in the USA over 3 years for a total of 11 replicates. We have found about 20 accessions that have moderate stripe rust APR with infection type (IT<5), and about 20 additional accessions that have novel seedling resistance (IT<3). We have genotyped the collection using genotyping by sequencing (GBS) and the 90K SNP chip array. GBS yielded a total 341K SNP that were filtered to 150K informative SNP. The 90K assay resulted in 11K informative SNP. We have conducted a genome-wide association scan (GWAS) and found one significant locus on 6BL ( -log p >5). Two novel loci were found for seedling resistance. Further investigation of the 6BL locus and the effect of *Yr36* showed that the 6BL locus and the *Yr36* have additive effect and that the presence of favorable alleles of both loci results in reduction of 2 grades in the IT score.

To identify alleles conferring adaption to extreme climatic conditions, we have associated the patterns of genomic variation in wild emmer with historic climate data from the accessions' collection sites. The analysis of population stratification revealed four genetically distinct groups of wild emmer accessions coinciding with their geographic distribution. Partitioning of genomic variance showed that geographic location and climate together explain 43% of SNPs among emmer accessions with 19% of SNPs affected by climatic factors. The top three bioclimatic factors driving SNP distribution were temperature seasonality, precipitation seasonality, and isothermality. Association mapping approaches revealed 57 SNPs associated with these bio-climatic variables. Out of 21 unique genomic regions controlling heading date variation, 10 (~50%) overlapped with SNPs showing significant association with at least one of the three bioclimatic variables. This result suggests that a substantial part of the genomic variation associated with local adaptation in wild emmer is driven by selection acting on loci regulating flowering.

Conclusions:

- Wild emmer can serve as a good source for novel APR and seedling R genes for stripe rust resistance.
= APR for stripe rust is a complex trait conferred by several loci that may have an additive effect.
- GWAS is feasible in the wild emmer population, however, its detection power is limited.
- A panel of wild emmer tagged with more than 150K SNP is available for further GWAS of important traits.
- The insights gained by the bioclimatic-gentic associations should be taken into consideration when planning conservation strategies.

**Description of Collaboration According to Goals**

**1) Development of a genetically characterized association panel of wild emmer for high resolution analysis of the genetic basis of complex traits;**

The origin of the panel's accessions is the seed bank of the Institute for Cereal Crops Improvement (ICCI) Tel Aviv university collection curated by Hanan Sela. He selected the panel out of the collection, propagated it, and sent DNA samples to Eduard Akhunov for genotyping with the genotyping-by-sequencing (GBS) platform and the 90K SNP chip array. Eduard Akhunov also conducted the bioinformatics analysis of the GBS and the QC of the 90K array.

**2) Characterization and mapping of genes and QTL for seedling and adult plant resistance to stripe rust in wild emmer populations;**

The panel was tested for stripe rust resistance at the seedling stage and at the adult-plant stage in the US by Brian Steffenson and in Israel by Hanan Sela. The results of the stripe rust resistance tests and the genotyping results provided by Eduard Akhonov were used by Hanan Sela to conduct genome-wide association scan (GWAS) to map QTL.

3) **Characterization of LD patterns along wild emmer chromosomes;**

The genotyping results provided by Eduard Akhonov were used by Hanan Sela to conduct linkage disequilibrium (LD) analysis.

4) **Elucidation of the multi-locus genetic structure of wild emmer populations and its correlation with geo-climatic variables at the collection sites.**

The genotyping results provided by Eduard Akhonov together with the georeference of the accessions and the heading dates provided by Hanan Sela were used by Eduard Akhonov to conduct the climate association analysis and the population structure analysis.

## Achievements

### Plant material

A diversity panel of wild emmer (480 accessions), representing the full geographic distribution of the species in Israel and covering ~120 collection sites, was selected from ~2,500 accessions from the Institute for Cereal Crops Improvement (ICCI) collection.

### Genotyping

The panel was genotyped with genotyping by sequencing method (GBS) (Poland et al., 2012) and by the 90K SNP Illumina chip array (Wang et al., 2014). GBS discovered 341,228 SNP. Two genetic matrices were generated from the raw GBS data for downstream analyses. 1) A matrix that was imputed using random forest machine learning approach (Poland et al., 2012) with 64K SNP and coverage of 100%. 2) Un-imputed matrix with at least 50% coverage of each SNP and each genotype and more than 1% minor allele frequency (MAF). This matrix consisted of 150K SNP. A third matrix was obtained from the 90K SNP array with 11,325 polymorphic markers (MAF>1%). Based on the genotypic data, we have conducted PCA and identified a separate group of ~70 accessions that belong to the subgroup of Judaicum. These accessions were removed from the analysis since our previous study showed that they have a large effect in elevating the LD that may cause false positives (Sela et al., 2014).

### Resistance Tests

The accessions were screened for seedling resistance by four PST isolates: one from Israel (#5006) and three from the USA (PSTv). Susceptible reactions (IT>6) were the most common phenotypes observed. For races PSTv-14,PSTv-37,PSTv-40, and #5006 resistance reactions were observed in 19,8,12, and 17% of the panel, respectively (Table 1). Only accessions susceptible to #5006 were included in the APR GWAS to avoid the masking of this trait by seedling resistance genes.

Field trials to estimate APR to PST were conducted in Israel at two locations: Bet Dagan in central Israel and at Barkai in northern Israel in the years 2016-18, and in the USA at UC Davis in the years 2016-17. In each of the trials, we planted 1-2 replicates (Table 2). We took readings of the infection types and recorded the growth stage. The infection types (IT) varied significantly between trials and years (Fig 1). The range of the ITs was 1-9 in all fields. We observed good correlation between replicates (Table 3). The repeatability (broad sense heritability, $H^2$) was 0.91. We have calculated the best linear unbiased predictor (BLUP) for the accessions using each replicate as an environment, and added the growth stage (Feekes) as a covariate (Endelman, 2011). The range of the BLUP values was 2.7-8.8, while the range of the BLUP for accessions that were susceptible in the seedling test was 4-8.8 (Fig 2). Twenty accessions with IT>6 at the seedling stage had BLUP IT values of less than 5 in the field. These accessions are considered to possess APR.

**Genome-wide Association Scan**

We used the R package rrBLUP (Endelman, 2011) to perform the genome-wide association scan (GWAS) with the three matrices. The GWAS was done using a mixed linear model (MLM) with relatedness matrix as random effects and 3-5 principal components of PCA generated from the SNP data as a Q matrix fixed effects to control for population structure. The environment (fields, replicates, years) as a cofactor, and the growth stage as a covariate were added to the model. GWAS was conducted for the whole panel or only for accessions with IT>6 at the seedling stage. The results are presented in Table 4. One locus on 6BL of three tightly linked GBS SNP was found to be significant in the APR GWAS of all the fields using the un-imputed matrix (-log p >5) (Fig 3, 4). The GWAS of the 90K SNP resulted in lower association values. The most significant loci were from 1B and 3B. The locus on 6BL had p values of less than 0.005 in six field sites and less than 0.001 in three field sites. As we did not observe any signal near the *Yr36* locus on 6BS, we have genotyped the accessions with *Yr36* positive marker (Fu et al., 2009). Accessions positive to *Yr36* had BLUP-IT values in the range of 4-8. The marker was found to have a significant association with IT (-log p >10) in a GLM analysis. Moreover, when testing association of the *Yr36* marker and the locos on 6BL with GLM, we found additive effect, meaning that genotypes positive for *Yr36* and the alternative allele of the 6BL locus have lower IT than *Yr36* alone (Fig 5). Several resistant accessions (IT>5) do not carry the favorable allele on 6BL and they probably carry other significant resistance conferring loci that were not detected, probably due to low frequency.

The most significant locus for seedling resistance was on chromosome arm 1BS near *Yr15* (Klymiuk et al., 2018). Removing accessions with alleles associated with seedling resistance on 1BS, revealed that additional loci loci from 5Al and 5BL were the most significant (Table 4).

**Identification of SNPs Contributing to Environmental Adaptation:** As a rich source of useful variation, the wild relatives of wheat have been broadly used in breeding. To identify alleles conferring adaption to extreme climatic conditions, we have associated the patterns of genomic variation in wild emmer with historic climate data from the accessions' collection sites. We genotyped a population of 475 geo-referenced wild emmer accessions from Israel using 90K iSelect SNP array and sequence-based genotyping resulting in a set of 37,983 SNPs mapped to the wild emmer genome (Avni e al., 2017). The analysis of population stratification revealed four genetically distinct groups of wild emmer accessions coinciding with their geographic distribution. Partitioning of genomic variance showed that geographic location and climate together explain 43% of SNPs among emmer accessions with 19% of SNPs affected by climatic factors. The top three bioclimatic factors driving SNP distribution were temperature seasonality, precipitation seasonality, and isothermality (Fick Stephen and Hijmans Robert, 2017). Integrated analyses of association between SNPs and these three bioclimatic factors using Bayenv (Coop et al., 2010) and association mapping approaches revealed 57 SNPs. We tested the overlap of climate-adaptive SNPs with genomic regions associated with contribution to heading date variation in wild

emmer. These regions were identified by performing genome-wide association study of heading date trait collected for the same panel of wild emmer at multiple year-by-location trials. Out of 21 unique genomic regions controlling heading date variation, 10 (~50%) overlapped with SNPs showing significant association with at least one of the three bioclimatic variables (temperature seasonality, precipitation seasonality, and isothermality). The identified loci were mapped to chromosomes 2A, 3A, 3B, 5A, 5B, 6A, 6B, and 7B, and two of them on chromosomes 5A and 5B were located in close proximity to vernalization gene loci *Vrn-A1* and *Vrn-B1*. This result suggests that substantial part of the genomic variation associated with local adaptation in wild emmer is driven by selection acting on loci regulating flowering, one of the major adaptive traits responsible for fine tuning the plant's development to local growth conditions. The manuscript presenting these results is currently being prepared for submission.

**Linkage Disequilibrium (LD) Patterns:** we have tested the extent of LD along the wild emmer chromosomes using either $r^2$ or LD p-values corrected for population structure. We found very rapid LD decay where $r^2$ values dropped to background levels within less than 100kb (Fig 6), while population corrected p-values dropped to background levels within less than 1kb.

**Discussion**

The wild emmer collection exhibited a large variation with respect to reactions to PST at the seedling stage and at the adult plant stage. Some of the variation could be attributed to known genes such as *Yr15* and *Yr36*. Nevertheless, we found resistant accessions at the seedling stage and at the adult-plant stage whose resistance could not be attributed solely to one of these genes. We were able to detect one putative locus that enhances *Yr36* resistance and several loci that may be associated with seedling resistance. However, we suspect that more causative genes are involved that could not be detected due to their small effect or low frequency that is below the detection power of GWAS. In the collection we found some degree of local adaptation to climatic factors and population structure associated with geography; the knowledge about the distribution of diversity according to these factors may help in planning collection strategies and selecting accessions for pre-breeding to drought tolerance.

**The Significance of Main Scientific Achievements:**

This is the first time that a large panel of wild emmer was screened specifically for APR. We found ~20 accessions that could be used as a source for stripe rust APR in breeding programs. The introgression of APR can be tracked by the significant QTL found in the GWAS. Additionally, we found novel sources of seedling resistance that could also be introgressed using markers from the seedling GWAS. An association panel of 480 wild emmer accessions tagged with >100K SNP markers is available for further GWAS of important agronomical traits.

Changes to original research Plan

We were planing to conduct the resistance tests in the US at Mt. Vernon, WA and at UC Davis, CA. The trial in Mt. Vernon did not succeed do to the whether that was not supportive to the growth of wild emmer wheat. Instead, we have extended the project for additional year and conduced two additional field trials in Israel in 2108 season.

## Publications for Project IS-4744-14CR

| Status | Type | Authors | Title | Journal | Vol:pg Year | Coun |
|---|---|---|---|---|---|---|

**Appendix**

**Table of Contents:**



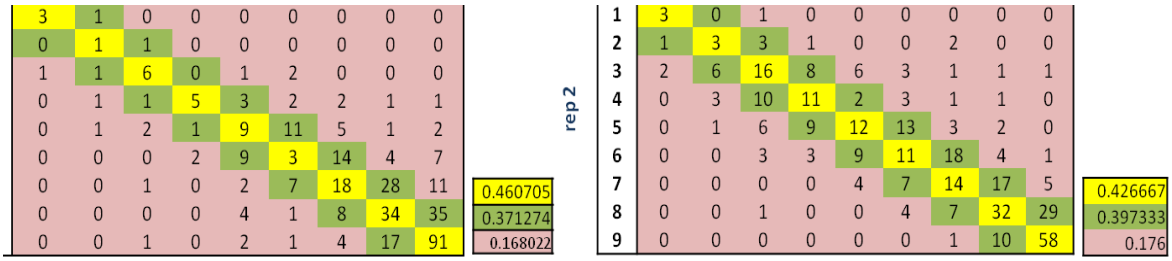
Table 1 Percentage of Resistant, Intermediate and Susceptible accessions at the seedling stage to three North-American and one Israeli PST isolates.

| Isolate | PSTv-14 | PSTv-37 | PSTv-40 | #5006 |
|---|---|---|---|---|
| R - IT= 0-3 | 19 | 8 | 12 | 17 |
| I - IT= 4-6 | 31 | 62 | 32 | 15 |
| S - IT = 7-9 | 49 | 28 | 55 | 67 |

Table 2 Field trials conduced in the project.

| **Location** | **Year** | **Number of replicates** |
|---|---|---|
| Beit Dagan, Israel 32.00N, 34.84E | 2016 | 1 |
| | 2017 | 2 |
| | 2018 | 1 |
| Barkai, Israel 32.47N, 35.02E | 2016 | 2 |
| | 2017 | 2 |
| | 2018 | 1 |
| Davis, California 38.53N, 121.72W | 2016 | 1 |
| | 2017 | 1 |
| Total | | 11 |

Table3. Correlations of IT within fields between two replicates. The small tables on the right indicate the proportion of accessions that were given the same IT score, One grade difference, or more than one grade difference scores respectively.



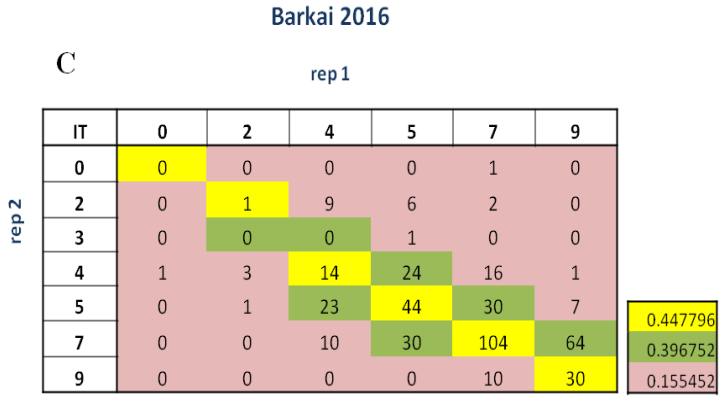




Table 4. Top Markers associated with stripe rust resistance. Analyses were done for APR, field resistance including accessions with seedling resistance, seedling resistance of genotypes with no *Yr15*. The GWAS was using either GBS data or 90K chip array data.

| **Analysis name** | **Marker name** | **Chr** | **Position** | **-log p value** |
|---|---|---|---|---|
| APR GBS | 6B_665186729 | 6B | 665186729 | 5.86 |
| APR GBS | 6B_665186737 | 6B | 665186737 | 5.86 |
| APR GBS | 6B_665186744 | 6B | 665186744 | 5.86 |
| APR 90k | TA006347-0547 | 1B | 57743619 | 3.76 |
| APR 90k | wsnp_BE498786B_Ta_2_1 | 3B | 87080095 | 3.68 |
| APR 90k | wsnp_Ex_c53976_57027635 | 1B | 525933671 | 3.43 |
| APR 90k | Tdurum_contig12926_816 | 5A | 41201100 | 3.19 |
| APR 90k | wsnp_Ku_c18023_27232712 | 5A | 664902198 | 3.06 |
| Field with seedling GBS | 1B_83299543 | 1B | 83299543 | 12.11 |
| Field with seedling GBS | 1B_83299570 | 1B | 83299570 | 12.11 |
| Field with seedling GBS | 1B_87982695 | 1B | 87982695 | 11.80 |
| Seedling GBS no Yr15 | 5A_674088833 | 5A | 674088833 | 7.03 |
| Seedling GBS no Yr15 | 5B_706615655 | 5B | 706615655 | 6.23 |
| Seedling 90k no Yr15 | BobWhite_c3194_125 | 6A | 622512532 | 4.05 |

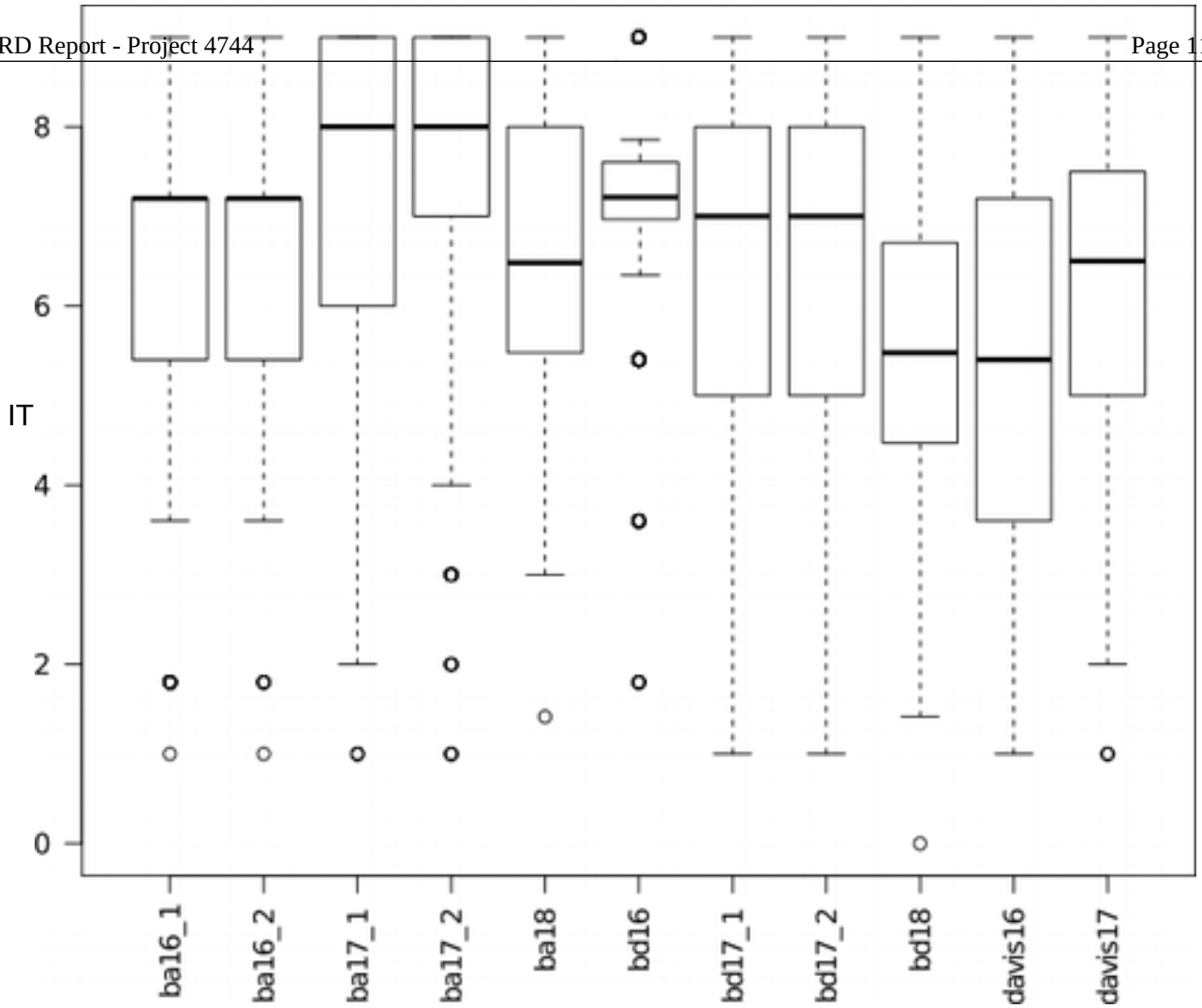



Fig 1 Distribution of IT in the field trials. Field names: ba= Barkai, bd=Biet Dagan.

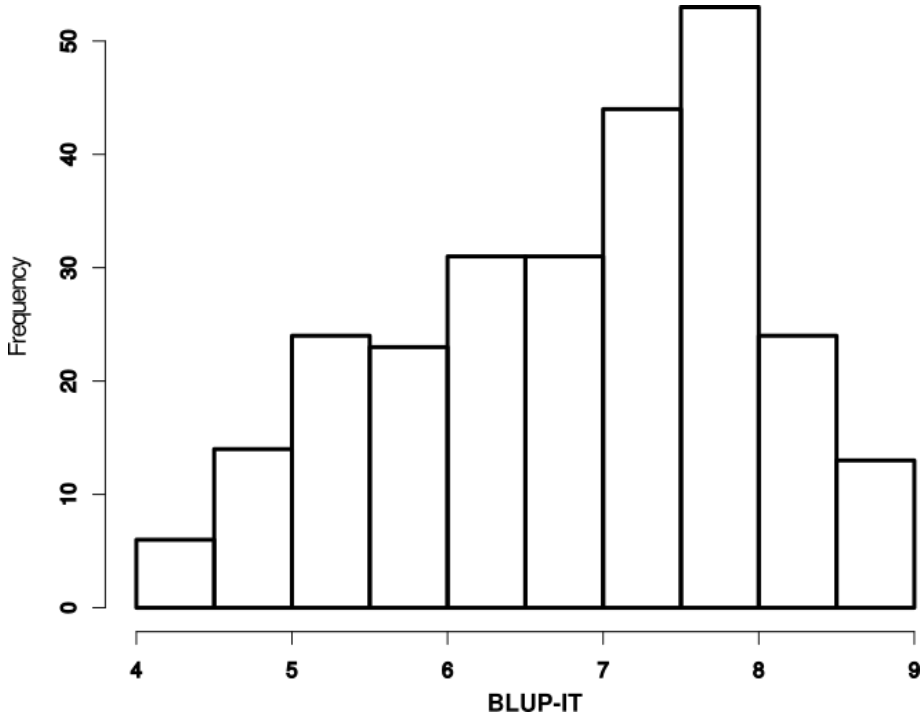



Fig 2 Distribution of BLUP-IT values for accessions with seedling resistance IT>6.

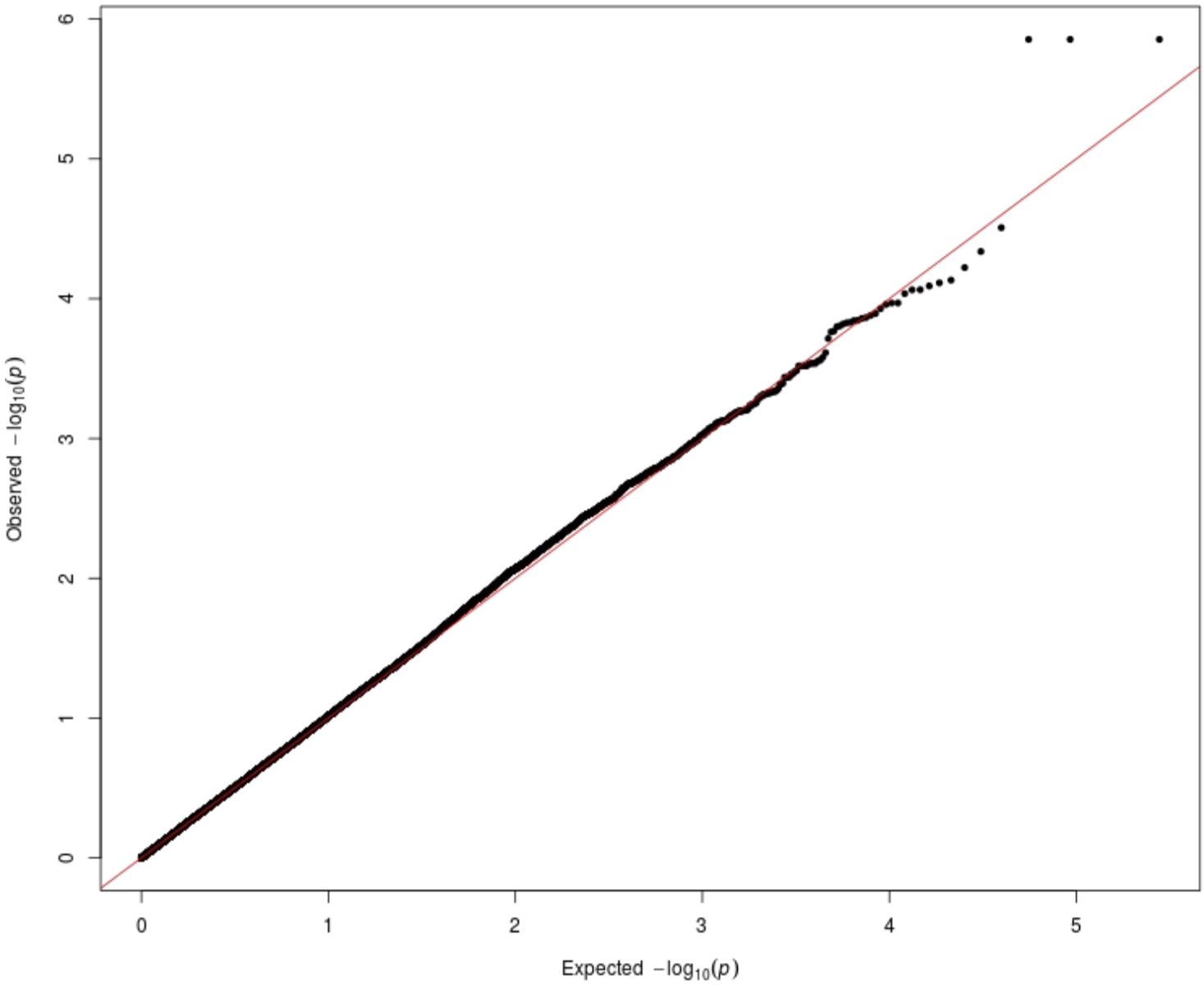



Fig 3. quantile- quntile (QQ) plot of p values from the GWAS of GBS generated SNP against adult-plant resistance IT scores.

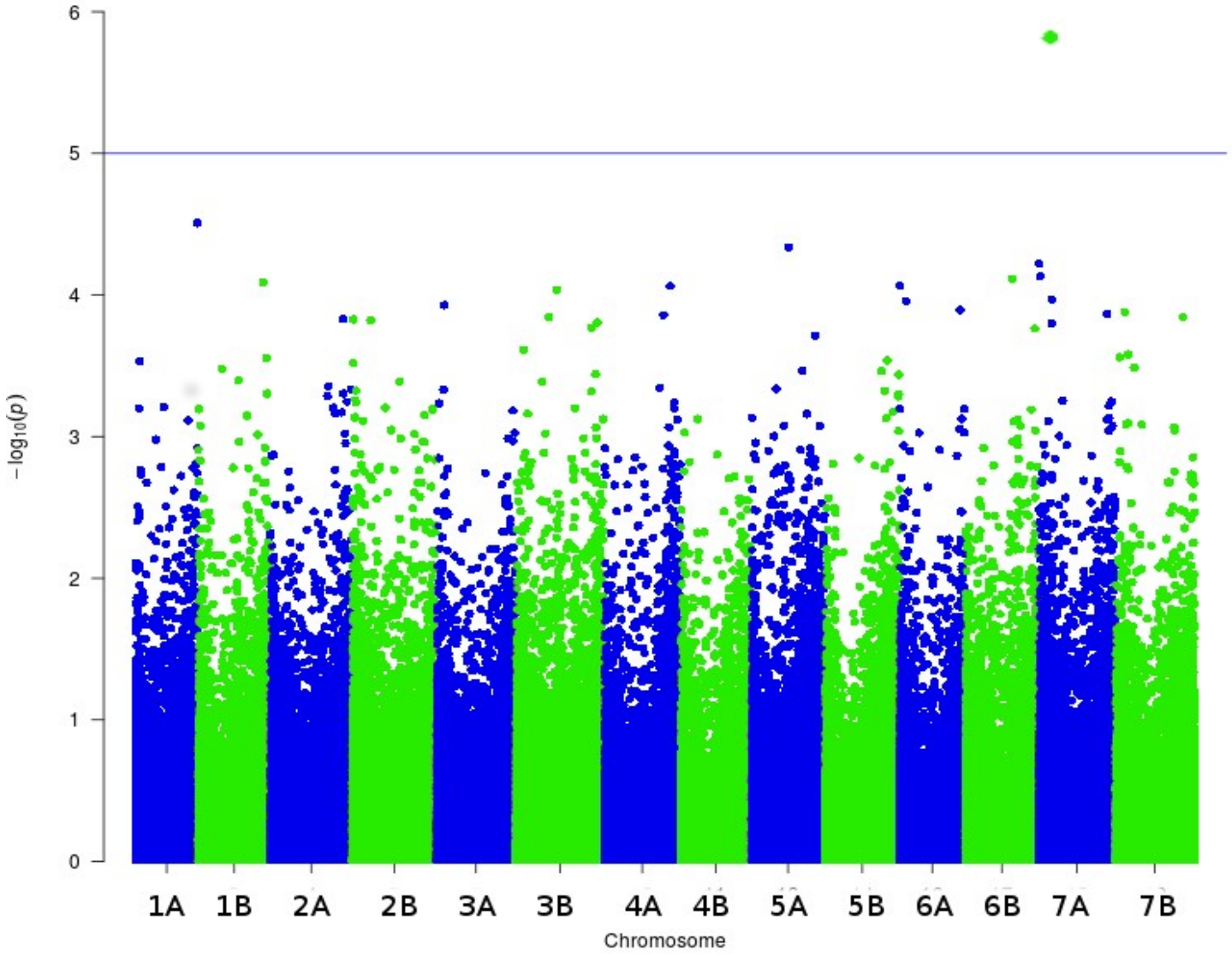



Fig 4. Manhattan plot of p values from the GWAS of GBS generated SNP against adult-plant resistance IT scores.

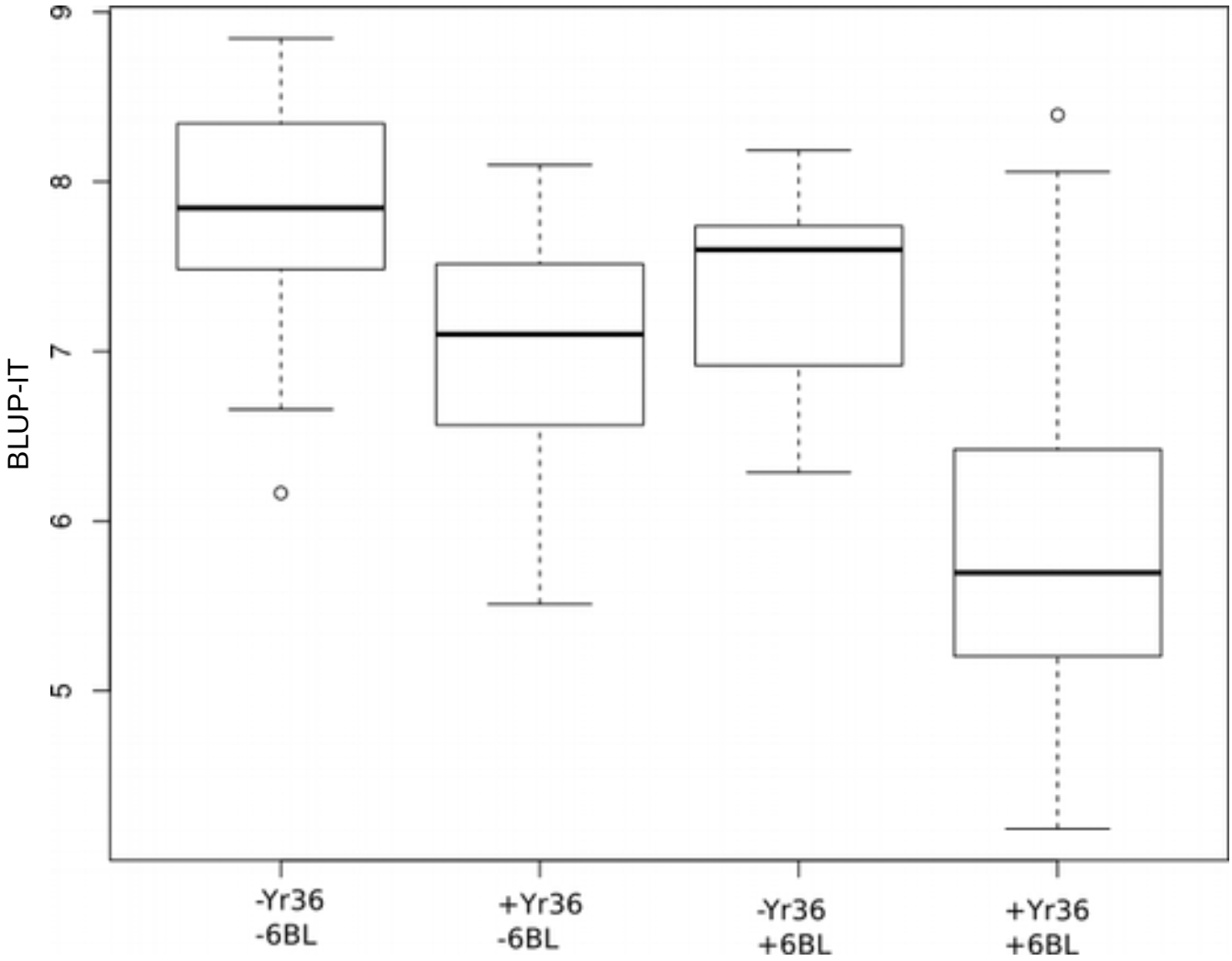



Fig 5. Box-plots of BLUP-IT of the four combinations of Yr36 and the locus on 6BL alleles. +6BL allele conferring resistance -6BL alternative allele. +Yr36 present, -Yr36 absence.

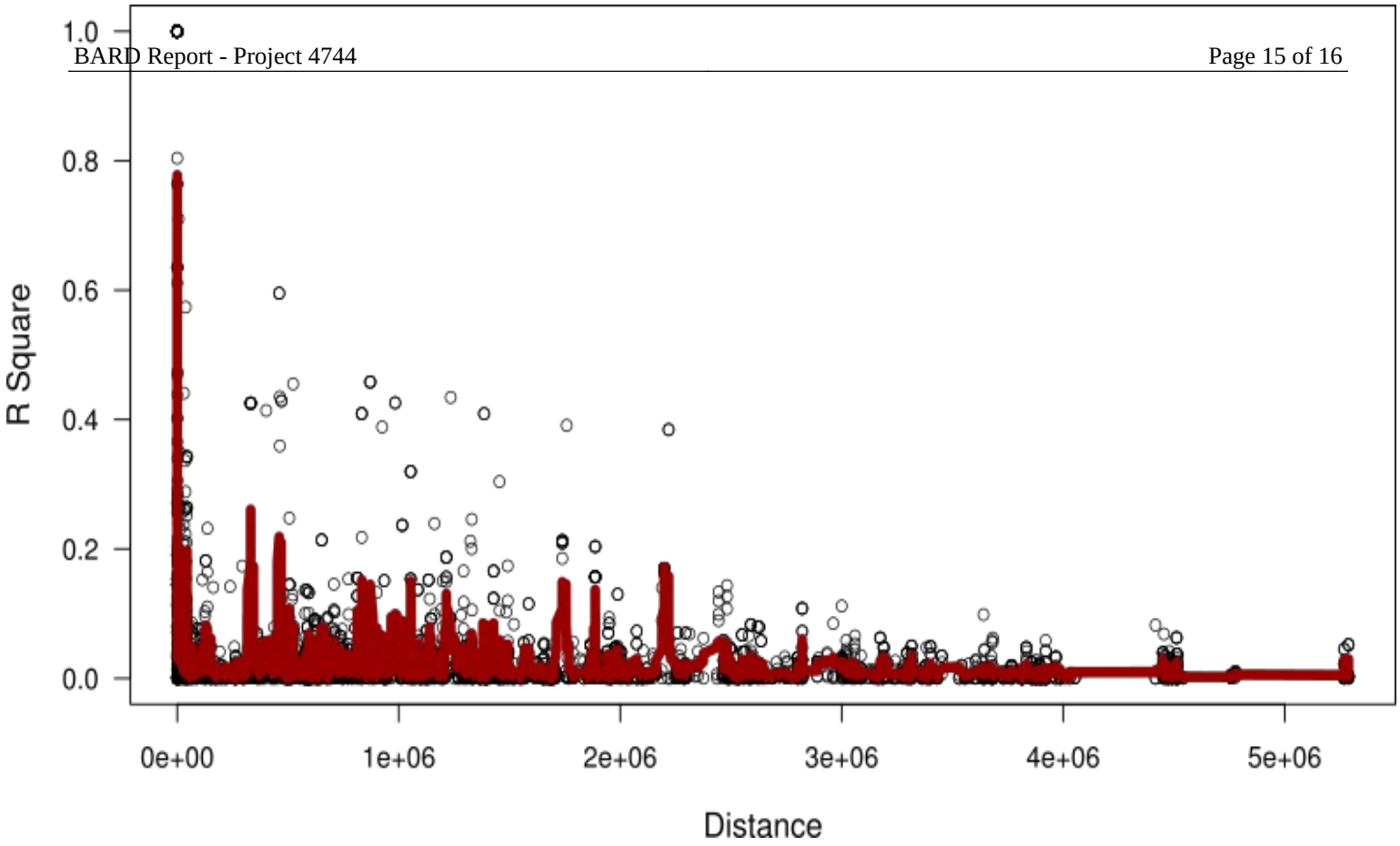



Fig 6. Linkage disequilibrium (LD) decay over distance (bp). LDs were calculated on sliding windows with 100 adjacent genetic markers. Each dot represents a pair of distances between two markers on the window and their squared correlation coefficient. The red line is the moving average of the 10 adjacent markers.

References:

Avni, Raz and Nave, Moran and Barad, Omer and Baruch, Kobi and Twardziok, Sven O and Gundlach, Heidrun and Hale, Iago and Mascher, Martin and Spannagl, Manuel and Wiebe, Krystalee and others. (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science. *357*, 93-97.

Coop, G., Witonsky, D., Di Rienzo, A. and Pritchard, J.K. (2010) Using environmental correlations to identify loci underlying local adaptation. Genetics. *185*, 1411-1423.

Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome. *4*, 250-255.

Fick Stephen, E. and Hijmans Robert, J.W. (2017) 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology.

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T. and Dubcovsky, J. (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science. *323*, 1357-1360.

Klymiuk, V., Yaniv, E., Huang, L., Raats, D., Fatiukha, A., Chen, S., Feng, L., Frenkel, Z., Krugman, T. and Lidzbarsky, G. (2018) Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase family. Nature Communications.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H. and Sorrells, M. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome. *5*, 103-113.

Sela, H., Ezrati, S., Ben-Yehuda, P., Manisterski, J., Akhunov, E., Dvorak, J., Breiman, A. and Korol, A. (2014) Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (triticum turgidum ssp. dicoccoides) population in Israel. Theoretical and Applied Genetics. *127*, 2453-2463.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G. and Cattivelli, L. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant biotechnology Journal.